DOCUMENT RESUME

ED 292 218 EC 200 751

AUTHOR Shaver, James P.; And Others
TITLE Is There A "Best" Way To Change Attitudes toward Persons with Disabilities? A Review of Research.
PUB DATE Apr 87
NOTE 61p.; Paper presented at the Annual Convention of the Council for Exceptional Children (65th, Chicago, IL, April 20-24, 1987). For a related document, see ED 286 335.
PUB TYPE Information Analyses (070) -- Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Attitude Change; *Disabilities; Effect Size; *Meta Analysis; *Research Methodology; Statistical Analysis

ABSTRACT

Meta-analysis was used to conduct an integrative review of the research on modifying attitudes toward persons with disabilities. Prior reviews, using small samples and lacking systematic data collection and analysis, were unable to draw firm conclusions about the effectiveness of attitude modification techniques. In the hope of correcting these deficiencies, an exhaustive search of all English-language research reports in the United States and Canada was undertaken that identified 667 potentially relevant items. Screening for relevance and adequacy of information reduced the accessible population for the integrative review to 273 reports describing a total of 644 treatment groups. A coding instrument containing some 162 categories was used and extensive quantitative analysis undertaken, but no clear-cut findings emerged. The major conclusion was that there had been a great deal of variety in the conditions under which the effectiveness of the various attitude modification techniques was investigated, that the variations were not systematically controlled, and that this confounded efforts to draw conclusions about treatment effectiveness. The consistent conclusion that research findings are equivocal may also reflect a reality that social-psychological phenomena are not amenable to systematization. Thirteen tables provide a detailed breakdown of study characteristics and effect sizes. (VW)

IS THERE A "BEST" WAY TO CHANGE ATTITUDES TOWARD PERSONS WITH DISABILITIES? A REVIEW OF RESEARCH*

James P. Shaver
Utah State University

Charles K. Curtis
University of British Columbia

Joseph Jesunathadas
Utah State University

Carol J. Strong
Utah State University

Legislation and judicial decisions are bringing handicapped persons into the mainstream of educational, social, and economic life in this society. Nevertheless, negative attitudes toward persons with disabilities continue to be detrimental to their potential to live dignified, productive lives and to contribute to society. A major research interest has been how to modify the negative attitudes and thereby mitigate the effects on persons with disabilities. That research literature has been reviewed in the past, but this paper is based on the most comprehensive review to-date (Shaver, Curtis, Jesunathadas, & Strong, 1987).

Prior Reviews of Research

Seven full and eight brief prior reviews of primary research on the modification of attitudes toward disabled persons were located. These reviews were examined for methodological soundness and for their contributions to knowledge using questions developed from the work of Jackson (1978, 1980) and others, with the primary research process as a model.

Although building on prior works is a standard approach for advancing knowledge in a field, most of the reviewers ignored previous, but relevant,

---

Paper prepared for a session at the annual convention of the Council for Exceptional Children, Chicago, April 22, 1987.

*This paper is based on a research project funded by the Research in Education of the Handicapped Program, Office of Special Education and Rehabilitative Services, U. S. Department of Education.

reviews. They did not draw on the findings of earlier reviewers; nor did they use inadequacies in prior reviews as a basis for improving the quality of their work. (This discussion is based on Shaver et al., 1987, Ch. 2.)

The methods used to locate and select primary studies were seldom reported, and the possibility of sampling bias was present in each review. The importance of sample selection can be illustrated by comparing the number of primary research reports cited in prior reviews with those identified for our review of literature. The total number of individual attitude change studies cited in the seven reviews and eight brief reviews was 192. The median number of primary studies referenced in the full reviews was 31 ($\bar{X}$ = 38; range, 24-70); in the brief reviews, the median was 11 ($\bar{X}$ = 13; range, 5-27). Our literature search yielded 273 primary research studies that met specific criteria for inclusion in our review of research on the modification of attitudes toward the disabled. An additional 394 studies were discarded as not suitable* for the present review. The limited reference lists and the small number of primary studies that were cited in more than one of the prior reviews cast serious doubt on the representativeness of the samples.

Many of the primary studies cited in prior reviews were low in treatment and internal validity; although this was mentioned in several reviews, it could not be determined how or if such studies were weighted when decisions concerning the effectiveness of particular interventions were reached. It seems apparent, given that lack of discussion, that treatment and internal validity were not explicitly considered in most reviews. Poorly designed and

---

*Of these, 363 were deemed irrelevant because they were, for example, correlational studies, used instruments that did not fit our definition of attitude, or attitudes toward mainstreaming rather than toward disabled persons were assessed. An additional 31 studies were discarded due to lack of information.

executed studies were included in the reviews without examining the association between design quality and outcomes.

Other methodological weaknesses were found in most of the reviews. Primary studies were placed into loosely defined intervention categories, with the result that important differences in sample and intervention characteristics were frequently disregarded. Narrative reports of programs and reviews of literature were cited as though they were primary studies. In several reviews, primary studies were misinterpreted and irrelevant studies were cited. Furthermore, there was a general tendency to report the findings of complex primary studies in simple treatment-outcome terms and, in some cases, to report only partial results. Moreover, even the statistical significance of findings was not presented in most reviews, and none reported research results in an effect size metric independent of sample size. And, studies which failed either to identify the dependent variable or to provide reliability or validity data for project developed instruments appeared to be accepted uncritically.

The most common conclusion in the reviews, as summarized in Table 1, was that there was not adequate evidence to support the effectiveness of any particular approach to attitude change. Although information plus direct contact with disabled persons seemed most likely to have positive effects, even those results were deemed equivocal because of limited samples, poor study designs, and inconsistent results. It was not clear whether the generally indefinite conclusions about the effectiveness of interventions for modifying attitudes accurately reflected the state of available research knowledge or were the result of the limited numbers of studies reviewed and the lack of a systematic approach to data collection and analysis. A

Table 1

Summary of Reviewers' Conclusions

Author	Contact	Information	Contact + Information	Vicarious/Simulation	Other Comments
Anthony (1971)	Studies with wide variety of disabled persons, no consistent changes (p. 119) Contact in and of itself does not change attitudes significantly (p. 120) May even reinforce neg. attitudes (p. 123)	Regardless of how info. presented, neglible affect (p. 120, 121)	Consistently favorable impact (p. 121, 123) Limited research, with college volunteers or trainees in helping professions: Dearth of data on other age groups, nonvolunteers, and nonhelping professions (p. 123)		Dearth of expt'l studies (p. 120) Need to include behavioral measures (p. 124) Little known about the time needed (varied in length from 6 hrs to 2 yrs, p. 123)
Haddle (1974)	No substantial results with contact alone (p. 93)	Most studies produced no significant results (p. 92)	Info. and contact tend to produce more significant results (p. 95) But studies poorly designed (p. 95) Most significant studies required extensive contact—ofter 40 hrs/wk (p. 95)		Cites Anthony (1972) that most Ss were volunteers and college age (p. 96) Most studies lacked good expt'l designs (p. 96)
Donaldson (1980)	Contact per se not effective (p. 505) Structured contact, pos. change (p. 505); unstructured social or prof'l contact, results equivocal (p. 505) Factors in pos. change: (1)=status (age; social, educ'l, vocational status; helping relation) (p. 505); (2)=disabled don't act in stereotyped manner (p. 507) Short, structered non-stereo. experiences, short term impact (p. 511)	No causal relationship between limited info. and attitude change (p. 508) If info. confirms negative stereotypes, negative affect (p. 511) Studies of courses not helpful because content unspecified and confounded with contact, media exposure, instructor characteristics (p. 508)		Simulation: only 2 studies. Can be effective if can observe reactions of nondisabled persons (p. 508)	Paucity of research: "literature contains relatively few studies" (p. 505) Failure to test theories (p. 529) Behavioral outcomes & long term effects need investigating (p. 512)

Table 1

Summary of Reviewers' Conclusions (continued)

Author	Contact	Information	Contact + Information	Vicarious/Simulation	Other Comments
Sandler & Robinson (1981)	Effects of contact equivocal (p. 98)	Effects assessed by few researchers (p. 99) Controlled studies needed (p. 100)	Cited 1 study that info and contact together beneficial (p. 101)		
Westwood et al. (1981)	"Results tend to be inconclu ive at best" (p. 221)	"Educational programs" produced equivocal results; results "inconclusive at best". Program content unknown (p. 221)		Simulation: "results . . .are inconsistent" (p. 222)	Contact: "earlier" studies didn't produce change (p. 221) Need to study various media (p. 221)
Towner (1984)	Various approaches with different populations equally effective; similar techniques with different disability groups yielded discouraging and contradictory findings. Positive and negative findings, in addition to non-significant results, from contact and information. Modes of presentation, including simulation, didn't produce significant differences (pp. 249-51)				Instrumentation seemed to have no effect (p. 251). Generally paper & pencil techniques (p. 224) Few attempts to address the complexity of attitudes (p. 224) Most reported no theoretical base (p. 224) Findings contaminated by methodological faults (p. 251)
Horne (1985)	Results inconclusive (pp. 156; 163-4) Limited # of studies, differences in treatments, methodology, & disabilities (p. 156-7)	Inconsistent results (pp. 163-4)	More successful, with prof'ls, but still not consistent (p. 151)	Role playing, children's books—results mixed (p. 178)	Interactions rarely assessed (p. 182, 186) Do immediate posttest results hold up? (p. 185)

Table 1

Summary of Reviewers' Conclusions (continued)

Author	Contact	Information	Contact + Information	Vicarious/Simulation	Other Comments
Fulton (1976)	Contact a factor but not with all social settings	Results with info. equivocal (p. 86)		Role play has potential (pp. 86-7)	Very few experiments that have positively changed attitudes toward physically stigmatized (p. 85)
Johannsen (1969)	Equivocal results (p. 224)	Not much is known about relative effectiveness of techniques (p. 224)			
Rabkin (1972)		Results conflict (p. 167)	Contact with patients and formal instruction effective (p. 166)		Questionnaires, few efforts to measure changes in behavior (p. 163)
Harth (1973)	Social contact not enough (p. 161)	More direct the procedure, the better the results (p. 160)	Effectiveness of knowledge through direct contact supported (p. 160)		No consistent line of research; no theoretical base (pp. 161-2)
Alexander & Strain (1978)*					
Segal (1978)	Can reinforce neg. attitudes if bizarre behavior (p. 215)		"Educated contact" necessary (pp. 215, 216)		
Horne (1979)			Need info. and contact (p. 63)		
Chuhon (1982)	Some indication that prof'l experience negatively related (p. 28)				Lack of definition of terms (p. 27) Methodology poor—lack of theory, standardized definitions, refined measurement devices Need to build on findings and experiences of other researchers (p. 27)

*No conclusions based on the research in regard to methods for modifying attitudes toward disabled persons could be found in the article.

comprehensive, systematic, meta-analytic type of study was undertaken to determine which was the case.

Procedures for this Review

Bangert-Drowns (1986) has noted that the choice of a quantitative approach for conducting an integrative review should be based on the purpose for the review. Our intent was to determine what the available research has to say about the effectiveness of treatments or interventions to modify attitudes toward persons with disabilities. For that reason, we adopted the approach to integrating the results of prior research that has been labeled by Glass (1976, 1977) as "meta-analysis". Properly implemented, the meta-analysis approach meets all of the criteria for high quality integrative reviews proposed by Jackson (1980). In conducting a meta-analysis, the reviewer: (1) locates either all studies or a representative sample of all studies on the defined topic; (2) converts the findings of each study, regardless of study quality, to a common metric--that is, computes an effect size for each relevant finding; (3) codes the various characteristics of each study that might have affected the results (such as type of treatment, methodological quality, sample attributes, and type of dependent measure); (4) uses statistics to summarize study outcomes (effect sizes) and to examine the covariations of outcomes and study characteristics; and, (5) draws conclusions based on the results of those analyses.

The Accessible Population of Studies

The purpose of this study was to conduct a comprehensive integrative review of the literature. The target population was all English-language reports of research identifiable through an exhaustive search conducted in

this country and Canada. There was no sampling procedure and only a few of the identified reports could not be obtained, although some that were relevant had to be discarded because adequate information was not reported. Therefore, the set of primary research reports that was reviewed was an accessible population, not a sample.

Of specific interest were empirical investigations of the effects of interventions, or treat ents, on the attitudes of nondisabled persons toward persons with disabilities. Correlational research was excluded. In addition to studies with experimental and quasi-experimental designs, single-group studies that involved a planned intervention and the collection of pretest and posttest data were included. Any research directed toward changing attitudes toward persons with disabilities or handicaps was of interest.

"Disabled or handicapped persons" was defined in terms of conventional special education categories, as reflected in Public Law 94-142, to include: mentally retarded, hard of hearing, deaf, speech impaired, visually handicapped, seriously emotionally disturbed (or, mentally ill), orthopedically impaired, deaf-blind, multi-handicapped, and learning disabled, as well as general categories such as "the disabled", "the handicapped", or "physically disabled". Studies of subjects from populations such as "disadvantaged students", "disruptive students", or "slow learners" were not included.

Attitudes toward disabled or handicapped persons was the dependent variable of interest in identifying and selecting primary reports. It was recognized that, consistent with common definitions (e.g., Triandas, Adamopoulas, & Brinberg, 1984), researchers might consider attitudes (which we defined, to provide context, as "interrelated beliefs about and feelings

toward an object which predispose the person to act in certain ways") as having cognitive, affective, and/or behavioral components. It was also recognized that "attitudes" might be assessed in a variety of ways, including paper-and-pencil tests with items that are cognitive-affective mixtures, assessments of changes in voluntary interactions with disabled persons, or reactions on projective-type tests. Measures which assessed only knowledge about the disabled did not qualify for selection, unless clearly considered by the research report author(s) to be attitude assessments; nor did measures which assessed attitudes toward mainstreaming qualify. General measures of attitudes toward children or other people were not included, unless specifically aimed at disabled persons or a particular type of disability, through instructions to the Ss or because of the context of the study--e.g., an attempt to change parents' attitudes toward their disabled children.

Measures such as sociometric scales, friendship choices, or observations of interactions were considered relevant only if clearly considered by the researcher(s) to be assessments of attitudes. Even if considered in the report to be attitude assessments, observational or other data were not included if the behaviors or responses of nondisabled Ss toward disabled persons, or the direction of behavioral or response change, could not be identified.

The search. The quest for research reports began with a computer search that included ERIC, CEC Abstracts, Dissertation Abstracts International, Index Medicus, Psychological Abstracts, and Social Science Research. The descriptor, "attitude change", was used with the broad descriptor, "disabilities", as well as with descriptors specific to types of disabilities such as "mental retardation" or "deaf". The computer search was updated

twice during the duration of the project. Hand searches of Psychological Abstracts, Education Index, and Dissertation Abstracts International were also done. Also, the references in Attitudes and Disability: An Annotated Bibliography, 1975-1981 (Regional Rehabilitation Research Institute on Attitudinal, Legal, and Leisure Barriers, George Washington University) were checked. In addition, the reference lists in all of the prior reviews cited earlier were searched, as was the reference list in each primary research report we obtained, whether or not it was decided to include the report in our review.

Copies of some 667 primary research reports that were judged potentially relevant based on title and abstract or reference in a review or primary research report were obtained through a variety of sources. The journal and the ERIC microfiche collections in the Utah State University, University of British Columbia, Simon Fraser University, and Western Washington University libraries were utilized. In addition, 218 requests for reports were sent by the Interlibrary Loan Department of the Utah State University library, of which 187 (86%) were received. Included were 77 dissertations, many of which had been identified in Dissertation Abstracts International. (No dissertation abstracts were included in the review because of the limited information they contain.) In addition, hard copies of 154 dissertations not available through Interlibrary Loan or from the authors were purchased from University Microfilms, Inc.

Each of the 667 primary research reports obtained was screened for relevance and adequacy of information. Letters were sent to authors requesting information when that in their reports was inadequate for effect size computations. One hundred and forty-six letters were sent for 117

reports. For 53 studies (45%), nothing was heard. For 13 reports, the letters were returned by the Post Office as undeliverable or someone wrote to say some such thing as that the author was dead or had moved leaving no forwarding address; for three reports, we were informed that the person to whom we wrote was not the author. For 23 reports (20%), authors wrote to tell us the information we had requested was not available. For 14 reports, information was sent that was different from that requested. Finally, for 14 reports (12%), we received information that allowed the desired effect size computations.

All told, 363 reports were discarded as irrelevant for our analysis and 31 were discarded for lack of information. (They are listed in the full research report: Shaver et al., 1987). The remaining 273 reports were the accessible population for the integrative review. (They are listed and a brief description of each study is presented in the full report, Shaver et al., 1987).

Instrumentation and Data Collection

The meta-analytic approach involves quantifying the outcomes of primary research studies using a common metric and coding various study characteristics so that it can be determined whether outcomes covary with the treatment variable and with any other study characteristics. The classification system used to code primary studies is, therefore, fundamental to data collection and data analysis. It must be comprehensive enough to "capture" the factors which are contributing to variance among studies, but not be so complex as to make coding overly burdensome. There are at least three other major considerations in developing a coding instrument: (1) That the data be collected in a usable format; (2) that the coding instrument

adequately reflect the substantive area under review; and, (3) that appropriate nontreatment study characteristics be coded.

In regard to format, a coding instrument developed at Utah State University's Early Intervention Research Institute for a meta-analysis of early intervention research with at-risk children (White & Casto, 1985) was of great value. Our prior review of research reviews helped to ensure that the second major consideration was met, as did the prior reading of a number of the primary research reports and tryouts of the instrument on research reports as it was developed. The basis for addressing the third major consideration was the literature on research design (e.g., Campbell & Stanley, 1963; Cook & Campbell, 1979; Shaver, 1983) and meta-analysis. Basic instrument development took place over a 3-month period; revisions continued until the scoring of new reports could be accomplished reliably, with no distortion of studies to fit the categories and no important information left out. An extensive set of conventions for coding studies was also developed.

The result of our instrument development was a coding instrument with some 162 categories, arranged in 10 sets according to the type of information to be coded, as follows: (1) General Information, such as date of publication and type of report (e.g., journal or dissertation); (2) Description of Sample, such as method of sample selection, sample size, percentage of males, educational level; (3) Treatment/Intervention, such as type of treatment (e.g., direct contact or information), the theory base, the treatment setting (e.g., classroom or mental institution), treatment characteristics (e.g., type of information and mode of information delivery), treatment verification efforts, and treatment validity; (4) Dependent Measures, such as type of measure, evidence on the reliability and validity

of scores; (5) Internal Validity, including various categories of threats, such as selection and history, and an overall rating on a three-point scale; (6) Results, including effect sizes; (7) Supplemental Information, such as whether the study was experimental or a program evaluation; (8) Prior Contact, including whether information about the subjects' prior contact with persons with disabilities was used in the analysis of data; (9) Contact (for studies of direct contact as an intervention), such as whether contact was voluntary and the relative status (e.g., education, age) of the persons involved; and, (10) Coding Summary, including who coded the study and how many minutes it took.

The quantification of results in a metric that is not relative to sample size--i.e., an effect size--is a major characteristic of meta-analytic research reviews. The major indicator of effect size for this study was Glass's Delta (Glass, McGaw, & Smith, 1981), which we labeled D. To compute a D, the difference between the experimental mean and the control group mean is divided by a standard deviation, if available, which is free of treatment effects. As our purpose was to obtain the most stable estimate of variance in the untreated population, we extended Glass's Delta by pooling the variances available for untreated groups--including treatment group pretest and control group pre- and posttest variances--to obtain the standard deviation by which the difference between means was standardized. When the means or the standard deviation for computing a D was not available, but the result from a test of significance, such as an F-ratio or t-ratio, was, D was estimated based on procedures spelled out in Glass et al. (1981).

Inter-rater Reliability

A rigorous criterion for reliability--90% agreement--was set, even though a criterion of 80% agreement is commonly used. The 90% criterion was

particularly stringent for inter-rater reliability because any categorization on which two or more of the three or four raters who were coding disagreed was coded as a disagreement.

Once adequate reliability was reached so that coding could begin, an inter-rater reliability check was conducted when any one of the raters had completed approximately 10 reports. Six separate reliability checks were completed; and for all but one, the 90% criterion was attained. For that one (85% agreement), a second study was coded, for which the criterion was met.

Because effect sizes are such a central part of a quantitative review, every effect size was re-checked for accuracy. Thirty-one errors were detected (and corrected), for an overall mean accuracy rate of 94%.

Intra-rater Reliability

After coding approximately 30 reports, each rater recoded one of the reports (selected by the project director) at the beginning of the sequence, without benefit of the first coding sheet. Again, the criterion was 90% agreement. Due to different rates of coding reports, one rater had three intra-reliability checks, one rater had two intra-reliability checks, one rater had one intra-reliability check, and one rater coded fewer than 30 reports so had no checks. All exceeded the 90% criterion.

Data Analysis

As Glass and his associates (1981, pp. 197-200) have pointed out, the role that statistical inference should play in meta-analyses is anything but clear. There was a major reason for not using inferential statistics in the integrative review reported here: the data to be analyzed constituted an accessible population, not a sample. The use of inferential statistics to

analyze data from an accessible population would be a perpetuation of ritual rather than a rationally justified procedure. Moreover, the use of an indicator of the significance of research results which is dependent upon sample size, as statistical probability is, is no more appropriate in analyzing the findings in an integrative review than it is in primary research (see, e.g., Carver, 1978; Shaver, 1985a, b).

In this study, the basic analytic approach was descriptive. Basic descriptive statistics were computed--means, modes, medians, standard deviations, and ranges. Two and three-way tables were used to investigate whether the treatment techniques and other characteristics of the studies in our accessible population were related to the size of effects.

The basic comparative base for an attitude change treatment was the absence of treatment--i.e., a control, placebo, or pretest condition*--rather than another treatment. When two treatment groups (i.e., Treatment A and B) were present in a study and each was compared with a control or placebo group, effect sizes were computed and coding conducted for the treatment versus control (T vs. C) or treatment versus placebo (T vs. P) comparisons, and not for the Treatment A versus Treatment B comparison. The basic data for analyses came from 644 T vs. C, T vs. P, and pre-post effect sizes.

Development of the coding instrument was guided by the admonition to include "all characteristics of the primary studies that are strongly suspected of affecting the findings . . ." (Jackson, 1978, p. 57). The upshot was a complex analysis process with difficult decisions about what to report and how. One major issue was how to handle data on the methodological quality of the studies in a data set.

---

*The single-group, pre-post design is, of course, a weak form of the control group design, with the pretest serving as an indication of attitudes in a no-treatment, control situation.

Quality of Research

The methodological quality of the studies from which effect sizes are collected has been a source of concern since Glass (1976) first proposed the use of the meta-analytic approach to integrative reviews. Although the concept of analyzing for the effects of study quality is still controversial (Bangert-Drowns, 1986), our stance in planning the procedures for this review was the same as Glass's: that is, include all studies, code for quality, and determine if effect sizes covary with study quality.

Quality Indicators

Although a number of our coding categories are related to quality of study, three global categories are particularly appropriate indicators of methodological soundness: general treatment validity, general internal validity, and adequacy of test validity. Each is widely regarded by researchers to be central to the validity of experimental results, and each is based on information from other categories.

Summary statistics for the three global indicators of quality are presented in Table 2. Two attributes of the data are striking: First, few studies received excellent or high ratings on any of the three types of global validity. Second, none of the ratings of validity explain much of the variability in effect sizes (as indicated by the Eta^2s of .01, .02, and .03). The low correlation between quality ratings and Ds is at least in part a function of the lack of variability in the former: Few effect sizes came from studies with excellent or high ratings.

To determine the association between Ds and membership in the higher frequency medium and low quality categories, point biserial coefficients were computed. The squared coefficients are .004, .01, and .001 for treatment,

Table 2

Quality of Study Indicators

General Treatment Validity

	Effect Sizes (Ds)		
Quality	N	Mean	SD
Excellent	4	.25	.30
Fair	245	.45	.68
Poor	395	.33	.56
Total	644	.37	.61

Note. Eta^2 = .01

General Internal Validity

	Effect Sizes (Ds)		
Level	N	Mean	SD
High	15	.89	.87
Medium	211	.32	.58
Low	418	.38	.61
Total	644	.37	.61

Note. Eta^2 = .02

Adequacy of Test Validity

	Effect Sizes (Ds)		
Adequacy	N	Mean	SD
High	9	1.13	.69
Moderate	520	.36	.62
Low	115	.40	.55
Total	644	.37	.61

Note. Eta^2 = .02

internal, and test validity, respectively, again indicating that very little variance in Ds was associated with quality ratings.

The lack of high quality in the research reviewed is probably due to two factors. The first is that attitude research is difficult to conduct, especially in applied settings (e.g., in elementary sch .ls) rather than laboratories. Another reason for the lack of high quality ratings is simply poor design and execution (as well as inadequate reporting, if better methodology was used than we were able to discern). Some examples from our data illustrate the point: For some 65% of the effect sizes, randomization of Ss was not reported. For only about 4% of the effect sizes were data collectors either fully or partially blinded. For 41% of the effect sizes, there was no mention of a reliability coefficient for scores on the dependent measure. And, for 83% of the effect sizes, the research reports contained no mention of any effort to verify implementation of the treatment independent variable. It would be difficult to argue that the available body of research on modifying attitudes toward persons with disabilities is exemplary in methodology.

The results with treatment, internal, and test validity posed a quandary. On the one hand, there appeared to be little association in our data set between the magnitude of Ds and the quality of the studies from which they come, at least as assessed via these global indicators. On the other hand, it can be argued (see, e.g., Bangert-Drowns, 1986, p. 392) that unless the studies being reviewed vary widely in methodological rigor, it makes little sense to examine study quality-outcome relationships. If this review had been conducted from a stance that studies with methodological flaws should be excluded from the analysis, our data set would have shrunk appreciably.

Some might even argue that we should not have attempted any integrative review. Slavin's (1986) proposal for "best evidence" research syntheses suggests otherwise. If high quality studies do not exist, it is appropriate to "cautiously examine the less well designed studies to see if there is adequate unbiased information to come to any conclusion" (p. 6). However, Slavin argues that a prior criteria should be applied in selecting "best evidence" studies, rather than quality-outcome analyses. We proceeded, then, with our analysis in a form of "best-evidence" review which Slavin did not intend to support. As Bangert-Drowns (1986) has pointed out, such a decision depends in large part on the purpose of the integrative review. An appropriate goal is to characterize the available research as a basis not only for insights into treatment effectiveness, but for decisions about further research. Careful summarization of the available past research is appropriate, even if only to make evident that which remains to be done.

Results

Some information from the analyses provides a context for consideration of the effect sizes for various treatments. For example (Table 3), there was nearly a balance between the number of comparisons for which the authors concluded their treatment was effective (N = 285; 44%) and those for which the treatment was deemed not to have had an effect (N = 259; 40%). Also, for 40 comparisons (6%), the report authors considered the results to be equivocal; and, for 19 comparisons (3%), it was concluded that the effect was negative. The actual number of effect sizes for which an attitude modification treatment group showed a negative change (that is, for which the treatment group's posttest mean was lower than its pretest mean) was 77 (12%), and 150 (23%) of the Ds were negative. It should not be easily

Table 3

Research Report Authors' Conclusions re Treatment Effectiveness

	Effect Sizes (Ds)			
Conclusion	N	%	Mean	SD
None stated	42	6	.34	.41
No effect	258	40	.03	.32
Equivocal	40	6	.51	.49
Produced effect	284	44	.74	.61
Negative effect	20	3	-.63	.36
Total	644	99[a]	.37	.61

Note. $Eta^2 = .37$.

[a]On this and later tables, percentages may not always add up to 100 because of rounding error.

assumed that the use of just any attitude modification technique will lead to a positive effect.

Comparisons of Experimental Treatments

What about the outcomes of the comparisons of experimental treatment groups against control or placebo groups or pretest scores? The various treatment techniques and combinations of techniques are briefly described in Table 4. They are arranged in rank order in Table 5, according to the magnitude of mean Ds. The mean effect sizes (Ds) for the attitude modification techniques can be viewed from two perspectives: (1) What does the average D for each treatment technique indicate about its effects as compared to no treatment? (2) What is indicated about the relative effectiveness of the different techniques?

Conventions to judge the magnitude of effect sizes must be used cautiously when the standards are arbitrary because there is no basis by which to judge the importance of variations in outcomes--as is the case with attitude assessments. It is, however, difficult to discuss results with no criteria in mind. Lacking more firmly grounded conventions, Cohen's (1977) criteria for small (d = .2), medium (d = .5), and large (d = .8) effect sizes provide a useful frame.

From that perspective, it is worth noting that none of the mean Ds reach the .8 criterion, although the mean D for the Persuasive Message studies is .67, closer to the large effect size criterion (.8) than to the medium one (.5). The differences between the Persuasive Messages mean D and the mean Ds for the other attitude modification techniques are all above an arbitrary standard for a trivial difference (.12--the magnitude of a difference between two Ds divided by the population standard deviation, .61, that would yield a

Table 4

Brief Descriptions of Attitude
Modification Techniques as Coded

Technique	Description
Information	Information on disabilities (e.g., etiology, characteristics, problems, similarities with nondisabled, prostheses) provided by means such as speakers, films, and books
Direct Contact	Ss in situation where they observe or interact with persons with disabilities
Vicarious Experience	Ss put in situations to help them experience what it is like to have disabilities
Persuasive Message	An argument presented via persons or printed or electronic media to convince Ss that they should have positive attitudes toward persons with disabilities
Persuasive Message, Contrast	Different messages or media used with treatment groups to investigate relative effectiveness
Systematic Desensitization	Thinking about disabled persons in relaxed, nonthreatening settings to extinguish negative attitudes
Positive Reinforcement	Use of classical or operant conditioning to modify behavior assumed to reflect attitudes
Other	Any combination of techniques other than Information Plus Direct Contact or Information Plus Vicarious Experience, which were coded separately

Table 5

Effect Sizes for Attitude Modification Techniques

Rank	Technique	Effect Sizes (Ds)			Differences Between Means[c]						
		N	Mean	SD	2	3	4	5	6	7	8
1	Persuasive Message	23	.67	.56	.16	.24	.27	.28	.35	.38	.47
2	Information Plus Contact	100	.51	.66		.08	.11	.12	.19	.22	.31
3	Direct Contact	93	.43	.73			.03	.04	.11	.14	.23
4	Vicarious Experience	58	.40	.76				.01	.08	.11	.20
5	Other	71	.39	.64					.07	.10	.19
6	Systematic Desensitization	21	.32	.44						.03	.12
7	Information	203	.29	.51							.09
8	Information Plus Vicarious	62	.20	.36							
	Persuasive Message, Contrast[a]	11	.13	.33							
	Positive Reinforcement[b]	2	(1.74)	(.01)							
	Total	644	.37	.61							

[a]Because ten of 11 Ds came from one study, the results are considered uninterpretable and the technique is not ranked.

[b]Too few effect sizes (less than 10) to be interpretable, and so not ranked.

[c]Numbers correspond to those for ranks of techniques. For example, the difference between the Persuasive Message mean (1) and the Information Plus Contact mean (2) is .16 (.67 - .51).

d = .2). Moreover, in three cases, the difference is greater than the standard for a medium difference (.31), approaching the standard for a large difference (.50) in one instance.

That messages developed purposely with an argument to sway attitudes would have the largest effect size, on the average, makes sense. It also may be of significance that 78% of the 23 Persuasive Message effect sizes come from studies in which the theory base (S-R/behavioral for 11, congruity/equilibrium for 6, and social judgment for 6) was explicit and the relationship of the theory to the treatment well-developed. (For "explicit theory base", the closest percentage was Systematic Desensitization with 76%, dropping then to Information Plus Vicarious Experience with 31%; for "explicit relationship to treatment", the same relationship held except that "Other" was third highest, with 34%.)

The Information Plus Contact studies produced the next largest mean D, .51, just over the arbitrary criterion for a medium effect size. Note again that the Information Plus Contact mean D is .16 below that for Persuasive Messages, barely larger than the arbitrary standard for trivial differences discussed above. At the same time, the differences between Information Plus Contact, on the one hand, and Direct Contact and Vicarious Experience, on the other (.08 and .11), are both less than the .12 trivial difference standard; but the difference for the Information Plus Contact mean D equals or exceeds the .12 criterion for all other comparisons, equaling the criterion for a moderate difference (.31) in one instance.

The next three mean Ds are clustered closely together--.43 for Contact, .40 for Vicarious Experiences, and .39 for Other (combinations of techniques other than the two in Table 4)--with Ds that fall at the midpoint

of Cohen's criteria for small and medium effect sizes (.2 and .5). The only difference between a mean D and one lower in the rankings that is non-trivial is between Other and Information Plus Vicarious Experience, a small difference (.19). The two remaining Ds--for Systematic Desensitization (.32) and Information (.29) are somewhat larger than the .20 small effect size standard, and only slightly higher than the means below them.

To sum up, although the mean Ds for the various techniques range from .67 to .20, clearly a broad range, there are no clear demarcations or groupings of techniques. In only one case (Persuasive Message versus Information Plus Contact) is the difference between contiguous means greater than our index of triviality (.12). The use of Persuasive Messages seems clearly to have resulted in larger Ds on the average than any other technique. Contact Plus Information runs a close second, and its use seems clearly to have produced larger Ds on the average than the use of Systematic Desensitization and the techniques ranked below it.

Treatment Variability--Heterogeneity of Ds

It might be tempting to look at the rankings in Table 5 as an index of effectiveness to be used in a singular fashion in selecting a technique to modify attitudes toward those with disabilities. That would, however, be too simplistic an interpretation of a complex set of data. To begin with, the standard deviations associated with each mean D serve as a reminder that the effects of each technique are not homogeneous; obviously, there is considerable overlap among the distributions of Ds for the various techniques. Moreover, it is important to remember that included in the Ds summarized by the means in Table 5 are negative values, indicating that, relative to the comparison group, a treatment had a negative rather than positive effect.

Table 6 presents a summary of the 150 negative effect sizes. Two things are worth noting: First, the percentage of negative effect sizes for each technique is roughly proportional to the percentage of effect sizes contributed to the total 644. No one technique contributed a markedly disproportionate number, or percentage, of negative Ds. But, second, it is remarkable that 23 percent (N = 150) of the 644 Ds were negative. Recall, too, that for 12% of the effect sizes, the treatment group had a negative change. Those figures not only highlight the need to keep variability in mind, but raise serious questions about the adequacy of the bases for the attitude modification treatments that were investigated. It also suggests that the treatments grouped under each technique label were not necessarily alike, even though quite different from those grouped under other labels.

Variation in Treatment Features

Heterogeneity in effect sizes may be due in part to the lack of homogeneity in treatment features within techniques. For example, in Tables 7 and 8, it can be noted that there was considerable variability in both the types of information and the modes of presenting it in the studies of the Information approach to attitude modification. The large number of Combination ratings for both types of information and delivery mode also suggest further variability, in the way that individual components were put together.

Variability within treatment categories is also evident for the Vicarious Experience and Persuasive Message studies (see Tables 9 and 10). It is interesting, as well, that while variations in type of information and mode of delivery accounted for about 6 to 7 percent of the variance in Information Ds, the percentage of variance attributable to treatment

Table 6

Negative Effect Sizes (Ds) for the
Attitude Modification Techniques

Technique	Negative Effect Sizes (Ds)				% of Negative Technique Ds[c]
	N[a]	%[b]	Mean	SD	
Persuasive Message	1/23	1/4	(-.36)[d]	(.00)[d]	(.04)[d]
Information Plus Contact	19/100	13/15	-.29	.29	19
Direct Contact	18/93	12/14	-.20	.17	19
Vicarious Experience	17/58	11/9	-.36	.42	29
Other	18/71	12/11	-.38	.31	25
Systematic Desensitization	4/21	3/3	(-.27)[d]	(.29)[d]	(19)[d]
Information	53/203	35/31	-.30	.32	26
Information Plus Vicarious	16/62	11/10	-.24	.19	26
Persuasive Message, Contrast	4/11	3/2	(-.14)[d]	(.10)[d]	(36)[d]
Positive Reinforcement	0/2	0/.3	—	—	—
Total	150/644	101/99.3	-.29	.30	23

[a]For N, the first figure is the number of negative effect sizes. The second figure is the total number of effect sizes.
[b]For %, the first figure is the percentage of the 150 negative effect sizes; the second figure is the percentage of the total 644 effect sizes.
[c]% of Negative Technique Ds is the percentage of the number of the Ds for a technique that were negative. E.g., 19% of the Information Plus Contact Ds were negative.
[d]Too few effect sizes (less than 10) to be interpretable.

Table 7

Types of Information Presented in Information Treatment Technique Studies

Information	Effect Sizes (Ds)		
	N	Mean	SD
Characteristics of disabled persons	2	(.52)[a]	(.21)[a]
Problems of being disabled	1	(-.03)[a]	(.00)[a]
Similarities with nondisabled	11	.11	.60
Managing disabled children	1	(-.17)[a]	(.00)[a]
How nondisabled react	11	.39	.20
How to relate in social situations	2	(-.77)[a]	(.28)[a]
Other	15	.22	.51
Combination	160	.31	.52
Total	203	.29	.51

Note. $Eta^2 = .06$.

[a]Too few effect sizes (less than 10) to be interpretable.

Table 8

Information Delivery Modes Used in the Information Treatment Technique Studies

Delivery Mode	Effect Sizes (Ds)		
	N	Mean	SD
Lecture	7	(.55)	(.71)
Discussion	11	.18	.27
Lecture–discussion	3	(.02)	(.25)
Print	23	.22	.44
Panel-disabled	1	(.80)	(.00)
Speaker-disabled	4	(.22)	(.17)
Film, video	21	.40	.58
Picture, filmstrip	4	(-.02)	(.63)
Audio	7	(.74)	(.59)
Simulations	1	(-.08)	(.00)
Regular course	24	.32	.74
Regular program	23	.18	.44
Other	7	(.27)	(.37)
Combination	67	.27	.46
Total	203	.29	.51

Note. For mean Ds and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

Eta^2 = .07.

Table 9

Types of Experience in Vicarious Experience Treatment Technique Studies

Experience	Effect Sizes (Ds)		
	N	Mean	SD
Role play	7	(.34)	(.43)
Simulation	26	.60	.72
Observe role play or simulation	2	(-.95)	(.46)
Video, films	9	(.05)	(.22)
Print, fiction or biography	2	(.05)	(.07)
Other	1	(-.09)	(.00)
Combination	11	.59	1.06
Total	58	.40	.76

Note. For mean Ds and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .20$.

Table 10

Types of Persuasive Messages Presentations in Persuasive Message Treatment Technique Studies

	Effect Sizes (Ds)		
Presentation	N[a]	Mean	SD
Video, film	3	(.52)	(.12)
Audio	3	(.31)	(.10)
Expert	8	(.48)	(.40)
Expert, disabled	1	(1.32)	(.00)
Other	8	(.99)	(.74)
Total	23	.67	.56

Note. $Eta^2 = .28$.

[a]All mean Ds and standard deviations are in parentheses because the number of effect sizes is less than 10 and too few to interpret.

variations is much larger for Vicarious Experience and Persuasive Message Ds--20% and 28%, respectively--suggesting that choice of technique features could be more important there.

There was also variability in the contact situations used in Direct Contact studies (see Table 11) and in the disabilities with which Ss were in contact (Table 12). About 12 percent ($Eta^2 = .12$) of the variance in Contact Ds was associated with situation differences, and about 10% with differences in disabilities. However, the Ns upon which most of the Ds in Tables 11 and 12 are based are so small as to make interpretation untenable. The lack of interpretability is compounded because two out of three Ds with sufficient Ns in Table 11 are for a Combination category and an amorphous "Other" category, the category in Table 12 with the largest N is "Combination", and only two other categories have more than 10 effect sizes in them. Consequently, while the data suggest diversity in the Contact studies, they tell us little about the effects of different types of content.

Attitudes toward . . .? An important treatment feature is the disability toward which the attitude modification efforts were directed. As Table 13 indicates, 44 percent (N = 286) of the effect sizes came from studies in which a target disability was not specified, but efforts were directed at changing attitudes toward an amorphous category of "disabled persons in general". The next most frequent change target, attitudes toward general physical disabilities (or, put differently, unspecified physical disabilities), was a distant second with 15 percent (N = 97) of the effect sizes. From there, the number of effect sizes for disability targets dropped off rapidly to 65 (10%) for Mentally Ill, to 37 and 36 (6% each) for Mentally Retarded--General (i.e., level of retardation not specified) and Combination

Table 11

Contact Situations for the
Direct Contact Treatment Technique Studies

Contact	Effect Sizes (Ds)		
	N	Mean	SD
As companion	8	(.42)	(.42)
As peer tutor	2	(.74)	(.29)
In cooperative learning group	3	(.22)	(.31)
As classmates	8	(1.13)	1.81
Practice teaching	4	(.23)	(.52)
In recreation program	4	(.52)	(.24)
Guest speaker	18	.24	.29
As teacher or counselor	8	(.13)	(.58)
Other	28	.50	.72
Combination	10	.33	.19
Total	93	.43	.73

Note. For mean Ds and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .12$.

Table 12

Characteristics of Disabled Persons in Contact Studies--Disabilities

Disability	Effect Sizes (Ds)		
	N	Mean	SD
Combination	20	.35	.52
Mentally Ill	15	.56	.53
MR--Mild/Moderate	14	.63	1.50
MR--General	6	(.51)	(.43)
MR--Can't Tell	5	(.23)	(.37)
Severe Multiple	5	(.36)	(.28)
Emotionally Disturbed	4	(.24)	(.27)
MR--Severe/Profound	3	(.50)	(.47)
Deaf	3	(-.16)	(.35)
Multiple Disabilities	2	(.67)	(1.61)
Physical--General	2	(.51)	(.03)
Wheelchair	2	(.42)	(.57)
Paraplegic	2	(.17)	(.18)
Blind	2	(.38)	(.35)
Hearing Impaired	1	(.75)	(.00)
Learning Disabled	1	(1.88)	(.00)
Can't Tell	6	(.18)	(.23)
Total	93	.43	.73

Note. For mean Ds and standard deviations in parentheses, the number of effect sizes is less than 10 and too few to interpret.

$Eta^2 = .10$.

Table 13

Disabilities Toward Which Modification Techniques Were Directed

Technique	Disabled General	Physical General	Mentally Ill	Mentally Retarded General	Combination	Hearing Impaired	Moderately Retarded	Severely Retarded	Visually Impaired	Other	Physically Impaired, Other	Emotionally Disturbed	Learning Disabled	Total
	Disability													
Persuasive Message	(.70)	(.49)	(.48)	(1.71)	—	—	—	—	—	—	—	—	—	.67
	(.05)	(.34)	(.40)	(.75)	—	—	—	—	—	—	—	—	—	.56
	3	9	8	3	—	—	—	—	—	—	—	—	—	23
Information Plus Contact	.53	(.66)	.20	(.65)	(.52)	—	(.91)	.50	(.71)	(.56)	—	(1.20)	—	.51
	.47	(.72)	.52	(.93)	(1.14)	—	(.96)	.32	(.23)	(.98)	—	(.25)	—	.66
	31*	9	22**	8	8	—	6	10	2	2	—	2	—	100
Direct Contact	.41	(.26)	.56	.20	(.74)	(.07)	(.91)	(.50)	—	(.29)	(.83)	(.24)	—	.43
	.59	(.26)	.53	.30	(.29)	(.53)	(1.84)	(.47)	—	(.31)	(.00)	(.27)	—	.73
	33	4	15	11	2	4	9	3	—	7	1	4	—	93
Vicarious Experience	.27	(.61)	(.41)	(.30)	(.79)	(1.47)	—	—	(.52)	—	(-.01)	—	—	.40
	.84	(.50)	(.30)	—	(.51)	(1.48)	—	—	(.27)	—	(.17)	—	—	.76
	29	7	4	1	4	3	—	—	3	—	7	—	—	58
Other	.41	.64	(.34)	(1.04)	(.37)	-.30	(.67)	—	—	(1.67)	(.04)	—	—	.40
	.47	.40	(.41)	(.88)	(.66)	.49	(.00)	—	—	(.23)	(.00)	—	—	.64
	30	11	4	6	3	13**	1	—	—	2	1	—	—	71
Systematic Desensitization	(.13)	(.30)	(-.10)	—	(.71)	—	—	—	(.25)	—	—	—	—	.32
	(.00)	(.55)	(.49)	—	(.20)	—	—	—	(.13)	—	—	—	—	.44
	1	5	4	—	6	—	—	—	5	—	—	—	—	21
Information	.23	.36	(.17)	(.19)	.44	(.92)	(.22)	(.12)	(.25)	(.42)	—	(.95)	(.18)	.29
	.51	.51	(.41)	(.28)	.71	(.59)	(.17)	(.41)	(.59)	(.18)	—	(.36)	(.30)	.51
	104**	43**	8*	6	13	4	4	7	8	2	—	2	2	203
Information Plus Vicarious Experience	.15	(.18)	—	(.17)	—	—	—	—	(.55)	—	(.10)	(.87)	(.59)	.20
	.35	(.18)	—	(.20)	—	—	—	—	(.21)	—	(.33)	(.00)	(.47)	.36
	44**	7	—	2	—	—	—	—	2	—	2	1	4	62
Persuasive Message, Contrast	.13	—	—	—	—	—	—	—	—	—	—	—	—	.13
	.33	—	—	—	—	—	—	—	—	—	—	—	—	.33
	11	—	—	—	—	—	—	—	—	—	—	—	—	11
Positive Reinforcement	—	(1.74)	—	—	—	—	—	—	—	—	—	—	—	1.74
	—	(.01)	—	—	—	—	—	—	—	—	—	—	—	.01
	—	2	—	—	—	—	—	—	—	—	—	—	—	2
Total	.29	.46	.31	.56	.55	.19	.76	.36	.37	.56	.09	(.68)	(.46)	.37
	.53	.52	.50	.75	.71	.92	1.[illegible]	.40	.42	.62	.30	(.49)	(.44)	.61
	286	97	65	37	38	24	2[illegible]	20	20	13	11	9	6	644

Note. The first number in each cell is the mean D, the second is the standard deviation, and the third the number of cases.

Means and standard deviations in parentheses are based on fewer than 10 cases.

*At least 10 fewer cases than expected, based on marginal frequencies.
**At least 10 more cases than expected, based on marginal frequencies.

(i.e., more than one disability target specified). Each of the other disability targets accounts for 4% or less of the 644 effect sizes.

Two types of treatment variability are evident in Table 13. First, the effects of each attitude modification approach have been investigated with several disability targets. Secondly, however, there is some clustering of disability targets within treatments. For example, Contact effect sizes have only come in substantial numbers (N of 10 or more) from studies directed at changing attitudes toward disabled persons in general, the mentally ill, and the mentally retarded in general.* Conversely, substantial numbers of effect sizes for the mentally ill as an attitude change target came from studies that investigated either Direct Contact or Information Plus Contact.

Moreover, not only the Ns but the effects are not consistent within disabilities or treatments. That is, there are differences in total mean Ds between disability targets (the Eta^2 for disability target and Ds is .05) and between mean Ds within disability categories as well. For example, there is a difference of .47 between the mean D for Disabled General (.29) and Moderately Retarded (.76); yet, within Disabled General, the range of mean Ds is from .15 (for Information Plus Vicarious Experience, ignoring the .13 for Persuasive Message, Contrast because 10 of the Ds came from the same study) to .53 (for Information Plus Contact), a difference of .38. By the same token, there is considerable variation in mean Ds within treatment categories. For example, for Information Plus Contact, the mean Ds range from .20 (Mentally Ill) to .66 (Physical General), a difference of .46. Although effect sizes might appear to be largely a function of treatment by

*It does not help interpretation that none of the target disability effect sizes for Persuasive Message as a technique, which had the highest overall mean D (.67), is based on an N of 10 or more.

disability attitude target interactions, the disparities in Ns for cells, as well as the large number of empty cells (not to mention the potential underlying interactions with other factors such as age of Ss), preclude such a conclusion--or even the use of analysis of variance to determine the proportion of the variance in Ds attributable to the treatment by disability interaction.

Other Study Characteristics

Are other study characteristics related to outcomes? What other reservations might be necessary in drawing conclusions about the results for different attitude modification techniques? For example, were effect sizes related to type of comparison--treatment versus control (T vs. C), treatment versus placebo (T vs. P), or single-group, pre-posttest (Pre-post)? Although the overall means for treatment versus control (T vs. C) and treatment versus placebo (T vs. P) comparisons (.36 and .29, respectively) were close to one another, the difference between each and the single-group, pre-posttest (Pre-post) mean D (.49) was .13 and .20, respectively. Yet, the Eta^2 for the relationship between comparison type and magnitude of D is only .01, reflecting in part the small numbers of T vs. P (N = 49; 7%) and Pre-post (N = 97; 15%) comparisons and the large number (N = 498; 77%) of the Ds in the T vs. C category. Moreover, the rankings and relative magnitudes of the mean Ds for the treatment technique remained essentially the same when the Pre-post means were excluded and the T vs. C and T vs. P means were pooled.

Pre-post comparisons did yield higher mean Ds for Information Plus Contact and Direct Contact, types of techniques likely to be used in college courses where pre-posttest data are often gathered (63 percent [N = 61] of the Pre-post Ds came from course and program evaluations, and nearly 60

percent of the course and program evaluation effect sizes came from samples of college and university students). So, even though pooling mean Ds from the three types of comparisons did not have a significant impact on the relative size or rankings of treatment technique means, it must be kept in mind that single-group, pre-posttest comparisons contributed heavily to the mean Ds for certain attitude modification techniques used with certain samples.

Time of Posttest

A small percentage (13%) of the 644 effect sizes came from follow-up posttesting. A pertinent question is, was the magnitude of Ds associated with time of posttest? There is little relationship between time of posttesting and D. A correlation between the number of weeks after the end of treatment when the posttest (immediate, delayed, or follow-up) was administered and the Ds for the 586 effect sizes for which that information was available yielded an r = .05--a very small relationship. No systematic time-of-posttest effect was observable among treatment techniques.

Type of Dependent Measure

Another aspect of testing that introduces study variability is the instruments used to assess attitudes. The assessment of attitudes in the primary research reviewed was dominated by questionnaires (N = 425; 66%) with Likert-type items. The next highest type of instrument, the semantic differential, was a distant second--N = 73; 11%. Only three other assessment types yielded data for at least 10 effect sizes: social distance scales (N = 58; 9%); adjective checklists (N = 32; 5%); and, "Other", a composite category of tests that didn't fit in any of the major categories, with 35

effect sizes (5%). The Eta^2 for Ds and type of assessment is .05. The association is not large, with assessment clearly dominated by Likert-type scales.

There were no systematic differences in the dependent measures used with different attitude modification techniques. That is, of course, due in part to the lack of variability in types of assessments—i.e., the prevalent use of questionnaires to assess attitudes. Perplexing questions of construct validity are raised by that use, with so little data coming from indirect, behavioral methods of assessment (e.g., Rokeach, 1968; Antonak, 1986). The perplexity is piqued by a low mean D (.16) for social distance scale assessments.

Length of Treatment

Length of treatment would seem to be an important study characteristic. Information on the total number of hours of treatment was available for 545 (84%) of the effect sizes in our data set. Length of treatment varied considerably, from .10 hour to over 1,000 hours. The mean number of hours of treatment was 37.14, with a standard deviation of 127.95. But the median number of treatment hours was only 4.00 and the mode was .7 hours—about the length of a typical class period.

For the 545 effect sizes for which the number of hours of experimental treatment was available, there was essentially no relationship between length of treatment and outcomes (r = .02). But there was an interaction between type of technique and length of treatment. For Information and Persuasive Messages, there were moderate negative associations (-.21, -.28), both of which, however, dropped to near zero (-.04, -.08) when outliers were excluded. There was a small negative r (-.20) for Information Plus Contact

when outliers were excluded (up from -.03). A coefficient of .60 for Systematic Desensitization makes sense, in that the effects of desensitization should increase with length of treatment, but the number of hours of treatment were clustered from 5 hours and less. For the other techniques, no relationship was evident.

Another source of treatment variability might be treatment context, the general milieu or environment within which a study was conducted. The effect sizes came largely from two contexts, college-university (49%; N = 314) and elementary-secondary schooling (36%; N = 235). Those two contexts accounted for 85 percent (N = 549) of the effect sizes. The overall means for the Elementary-Secondary Schooling and College-University contexts (.38 and .40, respectively) were remarkably similar.

Related to, but somewhat different from context is the setting of the research. While "context" refers to general environment, "setting" was defined as the specific type of place where the research was conducted. Forty-nine percent of the effect sizes (N = 314) came from studies carried out in regular classrooms, with 57 percent of those (N = 180) in public school classrooms and 35 percent (N = 109) in higher education classrooms. All of the 23 Ds from research in laboratory settings (15 of which were for Systematic Desensitization as a technique) came from the College-University Context, as did all but 5 of the 31 individual or small group setting Ds.

The size of the experimental treatment group is another variable of potential interest. Data on size were available for 642 effect sizes. The range in size was from 6 to over 800, but the modal group was relatively small, N = 20, and the median not much larger, N = 28. The correlation between Ds and experimental group Ns was practically nil for the total data

set (r = -.02). For individual treatment techniques with sufficient numbers of Ds to compute a coefficient, the r's ranged from essentially zero (Information [.03], Direct Contact [-.005], Information Plus Contact [-.11], Information Plus Vicarious Experience [.04], and Systematic Desensitization [.02]) to low and negative (Vicarious Experience [-.19], Persuasive Messages [-.20]), or low and positive (Other [.31]). In the case of the Other category, two outliers (high Ds and Ns) boosted the r. Overall, differences in treatment group Ns were not systematically related to the outcomes with different attitude techniques.

For the study characteristics discussed above (type of comparison, time of posttest, type of dependent measure, context, setting, and sample size), there appear to have been few systematically different effects for different treatments, and those that are discernible are often confounded with other sample attributes. Nevertheless, variability in study characteristics should be an ever-present consideration as our results are reviewed.

Sample Characteristics

Attributes of sample Ss might also influence outcomes. For example, the methods by which subjects were obtained might influence the nature of the sample. As might be expected, the samples for the studies which yielded our 644 effect sizes came from two major sources: volunteers (N = 220; 34%) and intact groups (N = 294; 46%), together accounting for 80 percent (N = 514) of the effect sizes. For the other 20%, random selection of individuals (N = 31) or of groups then used as the unit of analysis (N = 31) accounted for 62 effect sizes (10%); for 33 effect sizes (5%), selection was categorized as "Other"; and, for 35 effect sizes (5%), the method of sample selection could not be identified.

A mean D of .25 for Volunteers, as contrasted to a mean D of .42 for Intact Groups and .53 for Random Samples is somewhat perplexing in light of Rosenthal and Rosnow's (1975; Rosnow & Rosenthal, 1976) conclusion that volunteers are likely to be more intelligent, higher in need for social approval, and less authoritarian than nonvolunteers. However, as did Rosenthal and Rosnow (1975, e.g., p. 49), we included as volunteer subjects not only those who responded to a solicitation to participate in a research project but those who volunteered to participate in activities with disabled persons without knowing they were to be part of a research project. Such persons might have had even higher initial attitudes than would be expected on the general basis of estimated volunteers' traits, dampening treatment effects. However, the relationship of prior attitudes to outcomes could not be investigated with our data set because there were practically no reports of analyses in which pretreatment attitudes were included as an independent variable, with results reported by levels of antecedent attitudes.

Whether sample selection was related to attitude modification technique outcomes was difficult to ascertain because of the confounding of variables. Intact groups were more heavily represented than volunteers in Information Plus Contact, Direct Contact, and Information effect sizes, with the Volunteer mean D lower in each case; but 47 percent of the Information Plus Contact effect sizes, 49 percent of the Direct Contact effect sizes, and 46 percent of the Information effect sizes came from college and university samples, which are more likely to be obtained through solicitation of volunteers than are, for example, elementary and secondary school samples (40% of the college and university effect sizes came from volunteer samples, while only 22% of the elementary and secondary school effect sizes did; 57%

[N = 125] of the volunteer sample effect sizes [N = 220] were for college or university students and only 24% [N = 53] were for elementary or secondary school students).

The age of the Ss was a sample characteristic that we presumed might be related to treatment outcomes. Information on the schooling grade level of Ss was coded because it was more frequently reported and is a fairly close proxy for age, at least through the undergraduate years of college. As would be expected from the discussion of the contexts within which the attitude modification studies were conducted, 39 percent (N = 254) of the Ds came from studies conducted with Ss from preschool through high school and 43 percent coming from studies with undergraduates (N = 253) and graduate students (N = 29).

What about differences in grade-age level by treatments? Table 14 contains mean Ds for treatment techniques at each grade-age level. Purposely, only means have been included for which there were at least 10 effect sizes to make more graphic the pattern, including absences, of mean Ds. Information is clearly the most investigated technique, followed by Information Plus Contact. Just as clearly, the findings come primarily from Ss in the intermediate grades (many grade Combinations included intermediate Ss) and from undergraduates. Also, "adult" (Graduate, Postprofessional, and Adult Not In School) Ss received lower mean Ds than other Ss for Information and Information Plus Contact but not for Other (technique combinations other than those labeled in the table). And, the Direct Contact mean D for Intermediate Ss is strikingly higher than those for Combination and Undergraduate Ss, with the opposite result for Vicarious Experience--an Intermediate mean D (.25) .21 lower than that (.46) for Undergraduates. The

Table 14

Mean Ds for Treatment Techniques and Grade-Age Levels with At Least 10 Effect Sizes

Technique	Grade-Age Level											
	Preschool	Primary	Intermediate	Middle School	Junior High	Senior High	Combination	Undergraduate	Graduate	Postprofessional	Adult	Other
Persuasive Message												
Information Plus Contact			.59 15					.69 36	.09 12		.22 13	
Direct Contact			.31 16				.20 12	.40 44				
Vicarious Experience			.25 17					.46 25				
Other			.20 13					.32 26			.29 11	
Systematic Desensitization								.16 15				
Information		.32 14	.20 36	.20 10			.44 12	.37 79		.19 27		
Information Plus Vicarious Experience			.12 13				.21 22					
Persuasive Message, Contrast												
Positive Reinforcement												
Total N		14	110	10			46	225	12	27	24	468

Note. The first number in each cell is the mean D, the second number is N. Only means based on at least 10 effect sizes are included. Of the 644 effect sizes, 73% are represented in this table.

information in Table 14 suggests the necessary caution to avoid overgeneralizing the findings in this literature review to differing grade-age levels.

Prior reviewers (e.g., Horne, 1985, pp. 132, 143) have indicated that females tend to have more positive attitudes toward persons with disabilities, and that they may be more likely to change attitudes in a positive direction. In order to determine if gender was an important factor in study outcomes, we recorded the percentage of males in the experimental group. That information was available for 339 effect sizes. The mean and median percentages were nearly identical (35% and 36%, respectively); however, the mode was zero. That is, the most frequent occurrence was to have no males in the treatment group.

Overall, there was no relationship between percentage of males and outcomes ($r = .00$). The coefficients for the various treatment techniques ranged from moderately negative* ($r = -.47$, $r^2 = .22$ for Systematic Desensitization) to low and positive ($r = .31$, $r^2 = .09$ and $r = .27$, $r^2 = .07$ for Contact and Persuasive Messages, respectively), with most of the coefficients so low as to indicate negligible relationships.

When a complex analysis of variance was reported with gender and treatment as factors, Eta^2 was recorded an effect size for the interaction if it, or information to compute it, was available. Second, where Ds could be computed separately for males and females within a treatment by control, treatment by placebo, single-group, pre-posttest, or treatment A versus B comparison, we did so in order to analyze those Ds for differential treatment

*Recall that % of males was recorded, so a negative relationship indicates that with more males in the experimental group, Ds tended to be lower; conversely, a positive relationship indicates that with more males in the experimental group, Ds tended to be higher.

effects. The mean Eta^2 for the 36 available treatment by gender interactions was .02, with a standard deviation of .05. Separate Ds for males and females could be computed for 24 comparisons. The mean D for females was .41 and for males, .33 (the standard deviation was .49 for each), with a mean difference of .08 too small to be considered anything but trivial. The results are consistent with the overall r of .00 for percentages of males and Ds. Our analyses indicated little evidence that gender is consistently related to attitude change in our data set.

The extent and type of prior contact that Ss have had with persons with disabilities is another variable with potential power for mediating the effects of treatments. Although that factor was ignored by prior reviewers, we coded prior contact information. Assessment of prior contact was reported for only 260 out of 644 effect sizes (40%), with prior contact implicit for 29 effect sizes (4%)--e.g., inservice education with experienced psychiatric nurses or special education teachers. Those data did not yield much information because the relationship between prior attitudes and outcomes was hardly addressed. For only 4 effect sizes was there also a report of the correlation between prior contact and posttest attitude scores (N = 2) or prior contact and attitude change scores (N = 2), and for only 4 effect sizes was there a report of such correlations separately for a treatment and control group. Prior contact used with treatment as factors in a complex analysis of variance to determine prior contact-treatment interactions yielded only one effect size (Eta^2 = .02). In short, although prior contact is a potentially important variable, it received so little attention in our population of research reports that nothing can be said about the extent to which it might have mediated treatment effects.

The same is true of the personality characteristics of the treatment Ss. Personality attributes, such as authoritarianism, might well be important factors in the effects of attitude change efforts. However, only 3 effect sizes for personality by treatment interactions were identified in the population of studies (mean Eta^2 = .01).

Bangert-Drowns' (1986) portrayal of the general situation in summarizing psychological research provides an apt summary of the situation in regard to the variations in treatment and other study and sample characteristics as they might interact with interventions to modify attitudes toward persons with disabilities:

> Research outcomes vary in ways that make generalizable interpretations difficult. Such variation comes from a number of sources. It may reflect real population variation, the effects of different treatment features or study settings, sampling error, selection biases of the reviewer, publication biases, the effects of erroneous or insufficient reporting (unreported spurious influences, computational errors, typographical errors), differing degrees of validity and reliability in the outcome measures, and differences in the range or intensity of the independent variable. The task is enormous, but the power of social scientific inquiry would greatly increase if patterns could be found amid this outcome variation. (p. 396).

The patterns are not yet clear for the body of research we have reviewed.

What Is the Reality?

Prior reviews of the research on modifying attitudes toward persons with disabilities have not been based on comprehensive collections of research reports or on the systematic collection and analysis of extensive quantitative data on study outcomes and study characteristics. An assumption underlying this review was that the inability of prior reviewers to draw firm conclusions about the effectiveness of attitude modification techniques was likely due, at least in part, to the small samples of prior studies that were reviewed and the lack of systematic data collection and analysis. A meta-

analytic type of integrative review of the research on modifying attitudes toward persons with disabilities was proposed and initiated with the hope of bringing order to the literature where other reviews had not done so. As has been made clear above, that hope turned out to be in vain. Even with a population of studies based on an exhaustive search of the literature and with a quantitative integrative review technique, clear-cut indications were not found of the overall efficacy of techniques for modifying attitudes toward disabled persons or of reliable differences in efficacy between techniques.

As a consequence of the uneven distribution among treatments of variations in sample and other study characteristics, with many cells empty or with low Ns and the nesting of treatments, the analysis of potential concomitant variables was not particularly productive, except for indicating areas to be addressed in future research. Rather than drawing conclusions about the conditions under which different attitude modification techniques had been more or less successful, the major conclusion had to be that there had been a great deal of variety in the conditions under which the effectiveness of the various attitude modification techniques was investigated, that the variations have not been systematically controlled, and that, for that reason, they confounded efforts to draw conclusions about treatment effectiveness.

All possible data analyses could not be conducted within the time span of the funded project from which this paper has been prepared, and further analyses will be conducted for other reports to groups of professionals. However, at this time, the status of the research field might best be summarized with the flavor of a quote from Towner (1984) used earlier to indicate that another review of the literature was warranted:

> The applications [of similar techniques] yielded discouraging and contradictory findings. Both positive and negative attitudinal changes, in addition to numerous reports of [statistically] nonsignificant changes, resulted from interactions [of nondisabled persons] with disabled persons as well as from the provision of educational and general information. (p. 249)

The results of this review are likely to be disappointing for persons seeking guidelines for attitude modification programs.

This review indicates the need for both better designed research and a more productive research strategy, i.e., replication*, in the investigation of modifying attitudes toward persons with disabilities. However, the internal validity of attitude modification studies in this area is intrinsically frail because of the difficulties involved in studying such phenomena in applied settings. Even with careful replication, the accumulation of findings that indicate clearly what attitude modification techniques are most effective, or which are most effective with which types of persons for changing attitudes toward what types of disabilities, may turn out to be a difficult, if not impossible, goal to attain. That state of affairs may explain the results of this review and of the more limited reviews that preceded it.

Another possibility to be considered, however, is that the problematic state of research in this field is not a function of either poor design or inherent methodological deficiencies, but a reflection of reality. For example, Cronbach (1975) has argued that complex interactions among variables is the natural state of affairs with psychological phenomena. Perrow (1981), too, has contended that social-psychological phenomena may be much less amenable to systemization, and much more unpredictable, than most of those

---

*Only about 1.5% of the effect sizes come from efforts to replicate other studies.

engaged in "social science" research are willing to admit. And, based on their analyses of physical phenomena, physicists David Crutchfield and his associates (1986) have argued that "chaos"--that is, randomness generated according to orderly principles--may be a fundamental restriction on our ability to develop cause-and-effect conclusions in some areas of study. Contrary to the frequent assumption that the prediction of any phenomenon is possible, if only sufficient amounts of information can be gathered and processed, randomness, or chaos, and unpredictability may be fundamental with some phenomena.

Are attitudes and attitude change "chaotic" phenomena? That would explain the consistent conclusions in reviews of the literature that the research findings are equivocal. Improvements in research are necessary, as sketched out above, before it will be possible to know whether the observed state of affairs in research on modifying attitudes toward disabled persons is an artifact of methodology or a natural state of affairs.

How to modify attitudes toward persons with disabilities should be a continuing research agenda item. However, how to channel the behavior of individuals who have negative attitudes so as to avoid the restrictive and dehumanizing effects on persons with disabilities should continue to be both a policy and research focus. We ought not simply assume that research will someday tell us how to obliterate negative attitudes.

References

Alexander, C., & Strain, P. (1978). A review of educators' attitudes toward handicapped children and the concept of mainstreaming. Psychology in the Schools, 15, 390-396.

Anthony, W. (1972). Societal rehabilitation: Changing society's attitudes toward the physically and mentally disabled. Rehabilitation Psychology, 19, 117-126.

Antonak, Richard F. (1986, June). Methods to measure attitudes toward people who are disabled. Paper presented at the Conference on Attitudes Toward Persons with Disabilities, Hofstra University, Hofstra, New York.

Bangert-Drowns, Robert L. (1986). Review of developments in meta-analytic method. Psychological Bulletin, 99(3), 388-399.

Campbell, Donald T., & Stanley, Julian C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Carver, Ronald P. (1978). The case against statistical significance testing. Harvard Educational Review, 48(3), 378-399.

Chubon, R. (1982). An analysis of research dealing with the attitudes of professionals toward disability. Journal of Rehabilitation, 48(1), 25-29.

Cohen, Jacob. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cook, Thomas D., & Campbell, Donald T. (1979). Quasi-experimentation: Design & analysis issues for field settings. Chicago: Rand McNally.

Cronbach, Lee J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Crutchfield, James P., Farmer, J. Doyne, Packard, Norman H., & Shaw, Robert S. (1986). Chaos. Scientific American, 255(6), 46-57.

Donaldson, Joy. (1980). Changing attitudes toward handicapped persons: A review and analysis of research. Exceptional Children, 46, 504-513.

Glass, Gene V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5(10), 3-8.

Glass, Gene V. (1977). Integrating findings: The meta-analysis of research. Review of Research in Education, 5, 351-379.

Glass, Gene V, McGaw, B., & Smith, Mary Lee. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Haddle, H. (1974). The modification of attitudes toward disabled persons: The case for using systematic desensitization as an attitude-change strategy. American Foundation for the Blind Research Bulletin, No. 28.

Harth, R. (1973). Attitudes and mental retardation: Review of the literature. Training School Bulletin, 69, 150-164.

Horne, Marcia D. (1979). Attitudes and mainstreaming: A literature review for school psychologists. Psychology in the Schools, 16, 61-67.

Horne, Marcia D. (1985). Attitudes toward handicapped students: Professional, peer, and parent reactions. Hillsdale, NJ: Lawrence Erlbaum.

Jackson, Gregg B. (1978). Methods for reviewing and integrating research in the social sciences. Final Technical Report to National Science Foundation for Grant #DIS 76-20398. Washington, DC: The George Washington University, Social Research Group.

Jackson, Gregg B. (1980). Methods for integrative reviews. Review of Educational Research, 50, 438-460.

Johannsen, W. (1969). Attitudes toward mental patients: A review of empirical research. Mental Hygiene, 53, 218-228.

Perrow, Charles. (1981). Disintegrating social sciences. New York University Education Quarterly, 12(2), 2-9.

Pulton, T. (1976). Attitudes toward the physically disabled: A review and a suggestion for producing positive attitude change. Physiotherapy Canada, 28, 83-88.

Rabkin, J. (1972). Opinions about mental illness: A review of the literature. Psychological Bulletin, 77(3), 153-171.

Rokeach, Milton. (1968). Beliefs, attitudes, and values. San Francisco: Jossey-Bass, Inc.

Rosenthal, Robert, & Rosnow, Ralph L. (1975). The volunteer subject. New York: John Wiley and Sons.

Rosnow, Ralph L., & Rosenthal, Robert. (1976). The volunteer subject revisited. Australian Journal of Psychology, 28, 97-108.

Sandler, A., & Robinson, R. (1981). Public attitudes and community acceptance of mentally retarded persons: A review. Education and Training of the Mentally Retarded, 16, 97-103.

Segal, S. (1978). Attitudes toward the mentally ill: A review. Social Work, 23, 211-217.

Shaver, James P. (1983). The verification of independent variables in teaching methods research. Educational Researcher, 12(8), 2-9.

Shaver, James P. (1985). Chance and nonsense: A conversation about interpreting tests of statistical significance, Part 1. Phi Delta Kappan, 67(Sept.), 138-141. (a)

Shaver, James P. (1985). Chance and nonsense: A conversation about interpreting tests of statistical significance, Part 2. Phi Delta Kappan, 67(Sept.), 138-141. (b)

Shaver, James P., Curtis, Charles K., Jesunathadas, Joseph, Strong, Carol J. (1987). The modification of attitudes toward persons with handicaps: A comprehensive integrative review of research. Final Report to the U. S. Department of Education, Office of Special Education and Rehabilitative Services. Project No. 023CH50160, Grant No. G008530210. Logan: Utah State University, Bureau of Research Services.

Slavin, Robert E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational Researcher, 15(9), 5-11.

Towner, Arthurlene. (1984). Modifying attitudes toward the handicapped: A review of the literature and methodology. In R. Jones (Ed.), Attitudes and attitude change in special education: Theory and practice. Reston, VA: The Council for Exceptional Children.

Triandis, Harry C., Adamopoulos, John, & Brinberg, David. (1984). Perspectives and issues in the study of attitudes. In Reginald L. Jones (Ed.), Attitudes and attitude change in special education: Theory and practice. Reston, VA: The Council for Exceptional Children.

Westwood, M., Vargo, J., & Vargo, F. (1981). Methods for promoting attitude change toward and among physically disabled persons. Journal of Applied Rehabilitation Counseling, 12(4), 220-225.

White, Karl R., & Casto, Glendon. (1985). An integrative review of early intervention efficacy studies with at-risk children: Implications for the handicapped. Analysis and Intervention in Developmental Disabilities, 5, 7-31.